Misclassification of Smoking Status in the CARDIA Study: A Comparison of Self-report with Serum Cotinine Levels

Lynne E. Wagenknecht, DrPh, Gregory L. Burke, MD, Laura L. Perkins, PhD, Nancy J. Haley, PhD, and Gary D. Friedman, MD

ABSTRACT

Background. Although widely used in epidemiological studies, self-report has been shown to underestimate the prevalence of cigarette smoking in some populations.

Methods. In the CARDIA study, self-report of cigarette smoking was validated against a biochemical marker of nicotine uptake, serum cotinine.

Results. The prevalence of smoking was slightly lower when defined by self-report (30.9%) than when defined by cotinine levels equal to or greater than 14 ng/mL (32.2%, $P < .05$). The misclassification rate (proportion of reported nonsmokers with cotinine levels of at least 14 ng/mL) was 4.2% and was significantly higher among subjects who were Black, had a high school education or less, or were reported former smokers. Possible reasons for misclassification include reporting error, environmental tobacco smoke, and an inappropriate cutoff point for delineation of smoking status.

Using self-report as the gold standard, the cotinine cutoff points that maximized sensitivity and specificity were 14, 9, and 15 ng/mL for all, White, and Black subjects, respectively. The misclassification rate remained significantly higher in Black than in White subjects using these race-specific criteria.

Conclusions. Misclassification of cigarette smoking by self-report was low in these young adults; however, within certain race/education groups, self-report may underestimate smoking prevalence by up to 4%. (*Am J Public Health.* 1992;82: 33–36)

Introduction

Self-report is widely used to estimate the prevalence of cigarette smoking although it has been reported to underestimate the true prevalence in some populations.[1] Consequently, a number of biologic markers have been used to validate smoking, particularly in studies of cessation and of adolescent smoking.[2] Few studies have examined the validity of self-report in a general population sample of smokers and nonsmokers, nor in a sample of Blacks.

The purpose of this study is (1) to measure the discrepancy between smoking prevalence rates as defined by self-report and by serum cotinine, an objective measure of nicotine exposure, using a cotinine cutoff point of 14 ng/mL;[3,4] (2) to assess the degree of misclassification of reported nonsmokers by race, education, and past smoking behavior; and (3) to investigate possible reasons for misclassification.

Methods

The CARDIA study, a longitudinal epidemiological study of the risk factors for cardiovascular disease in a cohort of 5115 young adults aged 18 to 30 years, has been described previously.[5–7] The cohort was recruited to consist of approximately equal numbers of Black and White men and women of varied educational backgrounds. The data for this report were collected at the first clinical examination conducted in 1985 and 1986.

Smoking habit was assessed during the examination by interview. Smokers were defined as subjects who reported current, regular use of cigarettes (at least five cigarettes per week, almost every week). Nonsmokers were classified further as former smokers or as people who never smoked. Subjects who were reported being nonsmokers of cigarettes but who reported the current use of either cigars, pipe tobacco, smokeless tobacco, or nicotine gum were excluded from the analyses.

Smoking habit was also assessed at the recruitment telephone contact approximately 1 month prior to the examination. A smoker was defined as someone who smoked at least 100 cigarettes in his or her lifetime and who currently smoked. The reported number of hours per week that subjects were exposed to cigarette, cigar, or pipe smoke from others in their home was used as a measure of environmental tobacco smoke (ETS) exposure.

Serum cotinine was measured by radioimmunoassay[8,9] at the American Health Foundation in Valhalla, NY. Standard quality control procedures were followed; the interassay coefficient of variation was 7%. Based on previous investigations,[3,4] cotinine levels equal to or greater than 14 ng/mL were considered

Lynne E. Wagenknecht and Gregory L. Burke are in the Department of Public Health Sciences at the Bowman Gray School of Medicine in Winston-Salem, NC; Laura L. Perkins is with the School of Public Health at the University of Alabama at Birmingham; Nancy J. Haley is with the Metropolitan Life Insurance Company in Elmsford, NY; and Gary D. Friedman is with the Division of Research at Kaiser Permanente in Oakland, Calif.

Requests for reprints should be sent to Lynne E. Wagenknecht, DrPH, Department of Public Health Sciences, Bowman-Gray School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157–1063.

This paper was submitted to the journal December 26, 1990, and accepted with revisions May 28, 1991.

TABLE 1—Comparison of Smoking Prevalence Defined by Cotinine Versus Self-report, by Race and Education

Race/Education	n	Smoking Prevalence (%) Defined by Cotinine ≥ 14 ng/mL	Defined by Self-report	Difference	(95% CI)
Black ≤ HS	1317	47.1	43.8	3.3	(2.0,4.6)
Black > HS	1255	25.7	23.9	1.8	(0.8,2.9)
White ≤ HS	660	48.5	47.7	0.8	(–1.0,2.5)
White > HS	1752	19.6	19.8	–0.2	(–0.7,1.1)
Overall	4984[a]	32.2	30.9	1.3	(0.8,1.9)

Note. CI = confidence interval. ≤ HS = a high school education or less; > HS = more than a high school education.
[a]Thirty-five subjects were missing responses about cigarette smoking, 48 did not have cotinine results, and 48 reported use of other tobacco products, leaving 4984/5115 (97%) subjects available for analysis.

TABLE 2—Misclassification Rates, by Race and Education

Race/Education	All Reported Nonsmokers	Reported Never Smokers[c]	Reported Former Smokers
Black ≤ HS	8.2 (61/740)	5.1 (32/627)	25.7 (29/113)
Black > HS	3.7[aa] (35/955)	2.6[a] (22/841)	11.4[a] (13/114)
All Black	5.7 (96/1695)	3.7 (54/1468)	18.5 (42/227)
White ≤ HS	5.8 (20/345)	3.0 (7/230)	11.3[b] (13/115)
White > HS	2.1[aa,b] (29/1405)	1.0[a,b] (11/1092)	5.8[b] (18/313)
All White	2.8[bb] (49/1750)	1.4[bb] (18/1322)	7.2[bb] (31/428)
Overall	4.2 (145/3445)	2.6 (72/2790)	11.2 (73/655)

Note. Misclassification rate = proportion of reported nonsmokers with serum cotinine of at least 14 ng/mL. ≤ HS = a high school education or less; > HS = more than a high school education.
[a]$P < .05$, [aa]$P < .0001$ for difference in misclassification rates by education, within race.
[b]$P < .05$, [bb]$P < .0001$ for difference in misclassification rates by race, within education.
[c]All comparisons of misclassification rates by previous smoking history are significant, $P < .0001$.

TABLE 3—Distribution of Serum Cotinine Levels among Reported Nonsmokers and Smokers

Cotinine (ng/mL)	Reported Nonsmokers (n = 3445) n	Reported Nonsmokers (n = 3445) %[a]	Reported Smokers (n = 1539) n	Reported Smokers (n = 1539) %[a]
0–13	3300	—	78	—
14–49	72	49.7	133	9.1
50–99	23	15.9	142	9.7
100–149	15	10.3	206	14.1
150–199	7	4.8	197	13.5
200–249	8	5.5	220	15.1
250–299	9	6.2	151	10.3
300 +	11	7.6	412	28.2

[a]As a percent of all values of at least 14 ng/mL.

indicative of active smoking and not a result of ETS exposure.[10,11]

The Kolmogorov-Smirnov statistic was used to test for differences in the distributions of cotinine between Whites and Blacks and between former smokers and those who never smoked. Differences in ETS exposure between classes of nonsmokers were assessed using the Wilcoxon rank sum test. Confidence intervals were computed for differences between smoking prevalence rates using variance formulas for rate differences and taking the matched pairs into account.[12]

Results

Overall, the smoking prevalence rate as defined by cotinine levels of at least 14 ng/mL was 1.3 percentage points higher than that based on self-report (Table 1). The largest discrepancies were observed among the Black subjects; self-report underestimated smoking prevalence by 3.3 percentage points among Blacks with a high school education or less.

Of the 3445 reported nonsmokers, 145 had cotinine levels of at least 14 ng/mL, for a misclassification rate of 4.2% (Table 2). (The proportion of reported smokers with cotinine levels of less than 14 ng/mL was 78/1539, or 5.1%.) The misclassification rate was two times greater among Black subjects than among White subjects (5.7% vs 2.8%), almost three times greater among those with a high school education or less than among those with more than a high school education (7.5% vs 2.8%), and four times higher among the reported former smokers than among those who reported never having smoked (11.2% vs 2.6%, all $P < .0001$). Gender differences (not shown) were negligible.

The distribution of cotinine levels among the 145 subjects misclassified as nonsmokers differed from that among the true smokers, i.e., reported smokers with cotinine levels of at least 14 ng/mL (Table 3). One half of the misclassified subjects had cotinine levels between 14 and 49 ng/mL whereas less than 10% of the true smokers had cotinine levels in this range. The distribution of cotinine levels equal to or greater than 14 ng/mL was not significantly different either between former smokers and those who never smoked or between the races (not shown).

One half (73/144) of the subjects misclassified as nonsmokers reported spending most of their time around smokers whereas only 28% (906/3296) of the true nonsmokers (reported nonsmokers with cotinine of less than 14 ng/mL) reported this behavior ($P < .0001$). This was confirmed by their self-report of the number of hours they were exposed to ETS. The misclassified nonsmokers reported greater weekly exposure to ETS than the true nonsmokers (medians: 5 hours vs 0 hours, $P < .0001$); among the misclassified subjects, Blacks reported greater weekly exposure to ETS than Whites (medians: 6 hours vs 2 hours, $P = .02$). In addition, nearly one half (29/67) of the misclassified former smokers reported having quit smoking within the past year whereas only 20% of the true reported former

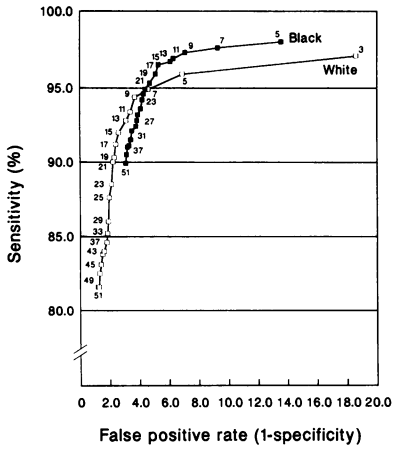

FIGURE 1—Receiver operating characteristic curve for varying levels of serum cotinine by race. The false positive rate (i.e., the proportion of reported nonsmokers whose cotinine level is equal to or greater than the given cutoff point) is the same as the misclassification rate as defined in this article. Levels of cotinine equal to or above the cutoff point are considered characteristic of a smoker. Self-report was considered truth for the calculation of these values.

TABLE 4—Sensitivity (%) and False Positive Rate (1-Specificity) for Various Cotinine Cutoff Values (ng/mL)

	Overall			Black			White		
Cotinine Cutoff	Sens.	False Pos. (1-Spec.)	Sum[a]	Sens.	False Pos. (1-Spec.)	Sum[a]	Sens.	False Pos. (1-Spec.)	Sum[a]
5	97.1	10.2	186.9	98.0	13.6	184.4	95.9	6.9	189.0
7	96.4	6.9	189.5	97.6	9.3	188.3	94.9	4.6	190.3
9	96.0	5.4	190.6	97.3	7.1	190.2	94.4	3.7	190.7[b]
11	95.4	4.9	190.5	96.9	6.3	190.6	93.4	3.4	190.0
13	95.0	4.6	190.4	96.7	6.1	190.6	92.8	3.1	189.7
14	94.9	4.2	190.7[b]	96.6	5.7	190.9	92.8	2.8	190.0
15	94.5	4.0	190.5	96.5	5.3	191.2[b]	92.0	2.6	189.4
17	93.9	3.7	190.2	95.9	5.1	190.8	91.2	2.4	188.8
19	93.2	3.5	189.7	95.3	4.7	190.6	90.3	2.3	188.0

Note. Cotinine levels < cutoff indicate nonsmoker, levels ≥ cutoff indicate smoker. Number of reported nonsmokers: 1695 Blacks, 1750 Whites; number of reported smokers: 877 Blacks, 662 Whites.
[a]Sum = sensitivity + specificity.
[b]Point at which sensitivity and specificity are maximized. Note that only odd values of cotinine are presented except for the value of 14 ng/mL.

smokers reported having quit over that period ($P < .0001$). Note that denominators are reduced slightly due to missing values.

Approximately one third (42/145) of the subjects misclassified as nonsmokers had identified themselves as current smokers at the recruitment interview one month earlier. Misclassification rates, calculated using smoking status as obtained at recruitment, were slightly higher than and yet consistent with those presented in Table 1; rates were twice as high in Black subjects than in White subjects (8.8% vs 3.8%) and three times higher in the reported former smokers than in those who reported having never smoked (14.5% vs 4.5%).

To investigate the possibility that the misclassification of these 145 reported nonsmokers was due to the use of an inappropriate cotinine cutoff point, we examined the receiver operating characteristic curve, which plots sensitivity against the false positive rate[13] (Figure 1). Cutoff points in the upper left corner of the graph maximize sensitivity and specificity. For these analyses, self-report was considered truth.

The cotinine cutoff point that led to the best combined levels of sensitivity and specificity in the overall cohort was 14 ng/mL (Table 4; note that sensitivity and specificity are both quite high over a range of cotinine cutoff points, from 9 ng/mL to 16 ng/mL). The race-specific cutoff point that maximized these parameters in the Black subjects was slightly higher than that in the White subjects (Figure 1, Table 4). (This was true across the entire range of sensitivity and specificity.) Among Black subjects, a cutoff point of 15 ng/mL maximized these parameters, yielding a sensitivity of 96.5% and a specificity of 94.7% (false positive rate of 5.3%). (Note that the definition of the false positive rate is the same as that of the misclassification rate). Among White subjects, a cutoff point of 9 ng/mL maximized these parameters, yielding a sensitivity of 94.4% and a specificity of 96.3% (false positive rate of 3.7%).

Applying these race-specific cutoff points to the entire cohort, the overall misclassification rate for reported nonsmokers was 4.5% (155/3445); these rates remained higher in Black subjects, as shown above (5.3% vs 3.7%, $P < .05$). Using these race-specific cutoff points to define smoking prevalence led to findings similar to those reported in Table 1, with a significant yet smaller underestimation of prevalence among Blacks (and also a significant underestimation among Whites with a high school education or less).

Discussion

These data indicate that self-report is a fairly accurate estimator of smoking prevalence in young adults. However, judging by cotinine levels and using a cutoff point of 14 ng/mL, smoking prevalence is underestimated to a greater degree among Blacks than among Whites, among former smokers than among those who never smoked, and among those with a high school education or less than among those with more.

Four possible reasons for misclassification were considered: (1) data collection/entry errors, (2) reporting errors, (3) ETS exposure, and (4) use of an inappropriate cutoff point for delineation of smoking status. In response to the first possibility, data collection followed standard quality control procedures. No data entry errors were found.

As for the second possibility, several types of reporting errors could have occurred. First, the subjects may have misunderstood the smoking questions. In fact, 29% of the 145 misclassified subjects identified themselves as smokers at recruitment. (It is possible that some of these subjects quit smoking in the month between the initial telephone contact and their CARDIA examination; however, we were unable to assess this.) Perhaps the questions asked at recruitment more readily identified the "light" smokers. Indeed, a large number of the misclassified subjects were light smokers, based on their cotinine levels. Considering this explanation, it was surprising that the misclassification rate was higher when comparing cotinine levels with smoking status reported at recruitment.

Differential reporting error could explain the higher misclassification rate among Black subjects. Compared with

Whites, Blacks may underreport smoking to a greater degree because they are, in general, lighter smokers.[6] However, when the smoking status reported at recruitment (the more liberal definition) was compared with serum cotinine levels, the difference in misclassification rates between White and Black subjects persisted.

Another explanation of misclassification, also a reporting error, is denial of smoking. Study subjects may wish to report behaviors consistent with a healthy life-style. Responses received in the possibly less-threatening recruitment telephone interview support this explanation. Denial may be an important factor particularly among the former smokers, many of whom reported having quit within the past year. A desire to "kick the habit" may lead these reported former smokers to underreport their habit to the apparent marked degree noted.

Furthermore, misclassification may partially be the result of an inappropriate cotinine cutoff point. However, this does not appear to be the case; 14 ng/mL was identified as the "best" cutoff point for the overall sample. This concern is more relevant to assessing the possible cause of higher misclassification rates observed among Black subjects. Race-specific cotinine cutoff points seemed necessary because of our earlier research, which showed significantly greater cotinine levels among Black smokers than among White ones over all levels of reported nicotine exposure.[7] Furthermore, the study that originally proposed the use of a cotinine cutoff point of 14 ng/mL was conducted with a small sample of White subjects and may not be generalizable to other samples.[4] However, the race-specific cutoff points that optimized sensitivity and specificity were not considerably different from 14 ng/mL, nor did they correct the higher misclassification rates in Blacks.

Our use of self-report as the gold standard for determining the best cotinine cutoff point seems incompatible with our initial challenge of the use of questionnaire data to assess smoking prevalence. Nevertheless, a measure of internal validity suggests that self-report is reasonably valid (i.e., the smoking status of 95% of the subjects was concordant in the tobacco interview and recruitment even though some discordance was expected).

Although, in general, exposure to ETS is an unlikely explanation for misclassification, it should not be dismissed in all cases. Reported ETS exposure was significantly higher among the misclassified subjects and among Blacks. This may partially explain higher rates of misclassification among Black subjects. Differences in nicotine metabolism (i.e., longer elimination times for nicotine and its metabolites even among nonsmokers) should also be considered in further investigations of this racial difference in cotinine levels.[7,14]

In summary, these data suggest that self-report is a reasonably accurate indicator of smoking prevalence in a biracial cohort of young adults. Misclassification is more pronounced among Black subjects, former smokers, and those with a high school education or less. Higher misclassification rates among Black subjects could not be explained by the choice of a cotinine cutoff point; however, reporting errors may account for higher misclassification rates among the former smokers and the less educated. Although we identified slightly different race-specific cutoff points than the previously proposed cutoff point of 14 ng/mL, these points had little effect on our findings. □

Acknowledgments

This study was supported by National Heart, Lung, and Blood Institute contracts N01–HC–48047, N01–HC–48048, N01–HC–48049, and N01–HC–48050. This paper was presented in part at the 30th Annual Conference on Cardiovascular Disease Epidemiology in San Diego in March 1990.

References

1. Coultas DB, Howard CA, Peake GT, Skipper BJ, Samet JM. Discrepancies between self-reported and validated cigarette smoking in a community survey of New Mexico Hispanics. *Am Rev Respir Dis.* 1988;137:810–814.
2. Stookey GK, Katz BP, Olson BL, Drook CA, Cohen SJ. Evaluation of biochemical validation measures in determination of smoking status. *J Dent Res.* 1987;66:1597–1601.
3. Cummings SR, Richard RJ. Optimum cutoff points for biochemical validation of smoking status. *Am J Public Health.* 1988;78:574–575.
4. Jarvis MJ, Tunstall-Pedoe H, Feyerabend C, Vesey C, Saloojee Y. Comparison of tests used to distinguish smokers from nonsmokers. *Am J Public Health.* 1987;77:1435–1438.
5. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol.* 1988;41:1105–1116.
6. Wagenknecht LE, Perkins LL, Cutter GR, et al. The impact of education on smoking in young adults: the CARDIA study. *Prev Med.* 1990;19:158–169.
7. Wagenknecht LE, Cutter GR, Haley NJ, et al. Racial differences in serum cotinine levels among smokers in the CARDIA study. *Am J Public Health.* 1990;80:1053–1056.
8. Haley NJ, Axelrad CM, Tilton KA. Validation of self-reported smoking behavior: biochemical analyses of cotinine and thiocyanate. *Am J Public Health.* 1983;73:1204–1207.
9. VanVunakis H, Gjika HB, Langone JJ. Method 16–radioimmunoassay for nicotine and cotinine. In: O'Neill IK, Brunnemann KD, Dodet B, Hoffmann D, eds. *Environmental Carcinogens Methods of Analysis and Exposure Measurement.* Vol. 9 *Passive Smoking.* 1987;317–330. IARC Scientific Publication, No. 81. Oxford University Press, NY.
10. Benowitz NL. The use of biologic fluid samples in assessing tobacco smoke consumption. In: Grabowshi J, Bell C, eds. *Measurement in the Analysis of Treatment of Smoking Behavior.* Rockville, Md: National Institute on Drug Abuse; 1983:6–26. NIDA Research Monograph No. 48.
11. Hoffmann D, Haley NJ, Adams JD, Brunnemann KD. Tobacco sidestream smoke: uptake by nonsmokers. *Prev Med.* 1984;13:608–617.
12. Rothman KJ. *Modern Epidemiology.* Boston, Mass: Little, Brown and Co; 1986:276.
13. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med.* 1978;8:283–298.
14. Sepkovic DW, Haley NJ, Hoffmann D. Elimination from the body of tobacco products by smokers and passive smokers. *JAMA.* 1986;256:863. Letter.